Osteoarthritis and Cartilage Open 7 (2025) 100556

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131

# Tibiofemoral cartilage strain and recovery following a 3-mile run measured using deep learning segmentation of bone and cartilage

Patrick X. Bradley [a], Sophia Y. Kim-Wang [b], Brooke S. Blaisdell [c], Alexie D. Riofrio [d], Amber T. Collins [b], Lauren N. Heckelman [b], Eziamaka C. Obunadike [b], Margaret R. Widmyer [b], Chinmay S. Paranjape [b], Bryan S. Crook [b], Nimit K. Lad [b], Edward G. Sutter [b], Brian P. Mann [a], Charles E. Spritzer [b,d], Louis E. DeFrate [a,b,e,*]

[a] Department of Mechanical Engineering and Materials Science, Duke University, United States
[b] Department of Orthopaedic Surgery, Duke University School of Medicine, United States
[c] Department of Physics, Duke University, United States
[d] Department of Radiology, Duke University School of Medicine, United States
[e] Department of Biomedical Engineering, Duke University, United States

ARTICLE INFO

Handling Editor: Professor H Madry

*Keywords:*
Auto-segmentation
Cartilage deformation
Cartilage thickness
Magnetic resonance imaging
UNet

ABSTRACT

*Objective:* We sought to measure the deformation of tibiofemoral cartilage immediately following a 3-mile treadmill run, as well as the recovery of cartilage thickness the following day. To enable these measurements, we developed and validated deep learning models to automate tibiofemoral cartilage and bone segmentation from double-echo steady-state magnetic resonance imaging (MRI) scans.

*Design:* Eight asymptomatic male participants arrived at 7 a.m., rested supine for 45 min, underwent pre-exercise MRI, ran 3 miles on a treadmill, and finally underwent post-exercise MRI. To assess whether cartilage recovered to its baseline thickness, participants returned the following morning at 7 a.m., rested supine for 45 min, and underwent a final MRI session. These images were used to generate 3D models of the tibia, femur, and cartilage surfaces at each time point. Site-specific tibial and femoral cartilage thicknesses were measured from each 3D model. To aid in these measurements, deep learning segmentation models were developed.

*Results:* All trained deep learning models demonstrated repeatability within 0.03 mm or approximately 1 % of cartilage thickness. The 3-mile run induced mean compressive strains of 5.4 % (95 % CI = 4.1 to 6.7) and 2.3 % (95 % CI = 0.6 to 4.0) for the tibial and femoral cartilage, respectively. Furthermore, both tibial and femoral cartilage thicknesses returned to within 1 % of baseline thickness the following day.

*Conclusions:* The 3-mile treadmill run induced a significant decrease in both tibial and femoral cartilage thickness; however, this was largely ameliorated the following morning.

## 1. Introduction

Advances in medical imaging techniques have enabled *in vivo* measurement of tissue biomechanical function and properties [1–3]. These insights are particularly critical to developing a greater understanding of joint health [4]. For example, there is significant interest in determining how activities such as running impact the function of articular cartilage [5,6]. To probe this question, magnetic resonance imaging (MRI)-based studies have investigated how acute bouts of running induce cartilage strain (i.e. change in cartilage thickness normalized to baseline thickness) within the tibiofemoral joint [7–9]. However, there is a paucity of data describing whether cartilage thickness recovers after running at time scales relevant to habitual loading, such as daily exercise [7]. Investigating cartilage thickness over time may provide insights into the influence of running on cartilage health.

*In vivo* measurement of articular cartilage thickness and strain often requires MRI-derived 3D models of the cartilage and bone. While some prior studies have utilized manual segmentation to create these models, this process can be time consuming [10,11]. In recent years, deep learning has demonstrated promise in automating some previously manual medical imaging segmentation tasks. One particularly successful deep learning segmentation architecture is the UNet, which several

* Corresponding author. Duke University, Box 3093, Durham, NC, 27710, United States.
*E-mail address:* Lou.DeFrate@duke.edu (L.E. DeFrate).

https://doi.org/10.1016/j.ocarto.2024.100556
Received 8 January 2024; Accepted 2 December 2024

groups have successfully implemented to automate the segmentation of bone, cartilage, ligament, and muscle from MRI scans [12–19]. Inspired by these prior works, we sought to develop deep learning models to automate the segmentation of the tibiofemoral bone and cartilage for our dataset of double-echo steady-state (DESS) MRI scans [7], in order to aid in the efficient measurement of cartilage thickness and strain.

Thus, the objectives of this work were to measure changes in cartilage thickness before and after a 3-mile treadmill run and to assess whether cartilage thickness recovered the following day. This was enabled via the development of deep learning models to automate segmentation of the tibiofemoral bone and cartilage from DESS knee MRI scans. Using our trained deep learning models to expedite segmentation, we hypothesized that both tibial and femoral cartilage would experience significant decreases in cartilage thickness following the 3-mile run and that these deformations would largely recover the next day.

## 2. Methods

### 2.1. Study design

Following institutional review board approval, eight asymptomatic males (age: 27–40 years; BMI: 18–25 kg/m$^2$) were recruited to participate in this study [7,20]. The study design consisted of an MRI-based protocol that began with participants arriving at approximately 7 a.m. and resting supine for 45 min to allow for consistent cartilage unloading before undergoing a pre-exercise DESS MRI scan [7,20]. All DESS MRI scans were collected on the same 3.0 T scanner (TrioTim; Siemens Healthcare; Malvern, Pennsylvania) with identical imaging parameters (field of view = 16 cm × 16 cm; image resolution = 0.3 × 0.3 × 1.0 mm; flip angle = 25°; repetition time = 17 ms; echo time = 6 ms). Following the first scan, participants performed a 3-mile treadmill run, followed immediately by a post-exercise DESS MRI scan. Participants returned the following morning at approximately 7 a.m., rested supine for 45 min, and underwent a recovery DESS MRI scan. This scan occurred an average of 22 h and 46 min (standard deviation = 5 min) after the end of the treadmill run the previous morning.

### 2.2. Segmentation model development & testing

In order to analyze the aforementioned data, we first sought to develop deep learning models to aid in the segmentation of the tibiofemoral bone and cartilage. All data utilized for model development were obtained from 6 previously published IRB-approved studies [21–26]. Data were aggregated from 2 of these studies to train the tibia and femur bone models, and from all 6 of these studies to train the tibial and femoral cartilage models. Table 1 presents the participant demographics for each aggregated dataset. The primary differences between the bone and cartilage datasets were that the cartilage datasets included data from more participants and participants with a more diverse injury status.

**Table 1**
Population variables used for model development.

| Variable | Bone (tibia/femur)[a] | Cartilage (tibial/femoral)[b] |
|---|---|---|
| Participants | 21 (13M/8F) | 72 (51M/21F) |
| Injury status | Healthy | Healthy (34)<br>ACL deficient (17)<br>ACL reconstructed (9)<br>ACL reconstructed & partial meniscectomy (12) |
| Age range (years) | 22–48 | 22–48 |
| BMI range (kg/m$^2$) | 20.0–27.9 | 18.5–34.7 |

[a] MRI data from the 21 participants was aggregated from published research [21,24].
[b] MRI data from the 72 participants was aggregated from published research [21–26].

Participants were not included if they exhibited high-grade chondral defects or full thickness cartilage loss. The inclusion of additional participants for the cartilage models was driven by the need to have more training data due to the increased variability of cartilage morphology compared to bone.

For each dataset involved in model development, all participants completed two DESS knee MRI scans, all of which were acquired using the same MRI scanner and imaging parameters described in the study design. MRI scans were previously segmented in prior works using commercially available solid modeling software Rhinoceros 4.0 (Robert McNeel and Associates; Seattle, Washington) by 6 researchers (E.C.O., M.R.W., C.S.P., B.S.C., N.K.L., E.G.S.) who were blinded to the order of scan acquisition when applicable. Each researcher received the same formalized task-specific training using a standardized training dataset, following standard guidelines, and was required to pass quality control testing administered by an expert radiologist prior to performing the segmentation task. The reviewing radiologist was fellowship-trained, currently with over 35 years of musculoskeletal (MSK) imaging experience and particular expertise in MRI and MSK biomechanics (C.E.S.). Subsequently, the same radiologist reviewed all segmentations of the study with the researcher. Furthermore, in each study, select data was resegmented to demonstrate reproducibility. Using custom MATLAB (Mathworks; Natick, Massachusetts) software, segmentations were converted into binary masks and automatically cropped from 512 × 512 pixels to 256 × 256 pixels surrounding the tissue of interest to aid in computational efficiency. All data were then randomly split into sets for training (cartilage: 100 scans/12,000 images; bone: 32 scans/3840 images), validation (cartilage: 22 scans/2640 images; bone: 6 scans/720 images), and testing (cartilage: 22 scans/2640 images; bone: 4 scans/480 images).

Both 2D- and 3D-UNet models were trained for each segmentation task in order to determine the optimal model architecture. All models were coded in Python 3.9 with Tensorflow 2.5 and Keras 2.5. Model training was performed using the Duke Compute Cluster, which features a variety of available GPUs. To optimize hyperparameter selection, a gridsearch for each model was performed over the following parameters: batch size, filter size, kernel size, and learning rate. Additionally, the 3D-UNet models included patch volume and stride length as parameters in the gridsearch with the following options implemented using the python module patchify: [volume: 256 × 256 × 112 pixels, stride: not applicable], [volume: 256 × 256 × 32 pixels, stride: 16], [volume: 256 × 256 × 16 pixels, stride: 8], [volume: 112 × 112 × 112 pixels, stride: 72], and [volume: 64 × 64 × 64 pixels, stride: 48]. Model depth was held constant at 5, and 2D and 3D models were trained for 500 and 2000 epochs, respectively, using early stopping with a patience set to 100 epochs. The Dice Similarity Coefficient (DSC) loss function was used for all models, which is equal to 1 – DSC and is a measure of the overlap between the manually segmented and model predicted masks [27]. Optimal models for each of the four tissues were determined by the model that achieved the highest validation set DSC during training. Subsequently, each of the four optimal models was applied to their corresponding testing set, and a new DSC was calculated.

Following model development and testing, the next objective was to evaluate the repeatability of using the trained bone segmentation models to create 3D models of the tibia and femur. To accomplish this, we utilized a separate dataset not involved in model development, consisting of seven new participants (sex: 3F/4M, age: 22–39 years, BMI: 19–28 kg/m$^2$) with no history of knee injury or surgery. For each participant, two MRI scans were acquired using the same imaging parameters as the data utilized for model development. Assuming that the tibia and femur bones should not deform significantly between the repeated scans, the objective was to use the trained bone segmentation models to create 3D models of the tibia and femur and measure bone surface distances between the repeated scans. These distances represent differences in how the trained segmentation models predicted the scans and can be used as a measure of scan-to-scan repeatability.

The pipeline for this process began by inputting the two DESS MRI scans of the same participant into the trained segmentation models, which generated masks for each of the bones (Fig. 1). Outlier pixels disconnected from the main bone regions of interest were manually removed and mask contours were extracted and converted into point clouds for all images stacked together. Point clouds were then reconstructed into 3D surface mesh models using Geomagic Studio 11 (3D Systems; Research Triangle Park, North Carolina), and the 3D models of each bone were aligned with one another [7]. Finally, surface distances were calculated in the x- (anterior-posterior), y- (superior-inferior), and z- (medial-lateral) directions at different regions of interest (within a 2.5 mm radius) sampled evenly across the femoral condyles and tibial plateaus. A total of 36 regions of interest were measured across the femoral condyles and 18 regions of interest were measured across the tibial plateaus [28]. Mean distances in the x-, y-, and z-directions for each participant were used to calculate a group mean and 95 % confidence interval for each measurement direction.

The next task was to assess the repeatability of cartilage thickness measurements by evaluating cartilage thickness from scans acquired on different days. Using the same dataset as was previously described in the study design, we used the pre-exercise and recovery scans, as these scans both represent unloaded time points when the cartilage is at baseline thickness. It should be noted that day-to-day fluctuations in cartilage thickness could impact measurement repeatability. In order to enable repeated site-specific cartilage thickness measurements, 3D surface mesh models of the associated tibia and femur from each scan were created and aligned together along with the corresponding cartilage surfaces (Fig. 2). Cartilage thickness was then measured by sampling and averaging 18 locations across the tibial plateaus and 36 locations across the femoral condyles (Fig. 1). At each location, all cartilage vertices within a 2.5 mm radius were included. Cartilage thickness at each vertex was defined as the distance to the nearest point on the bone surface. Finally, the thicknesses at all vertices were averaged together to produce a thickness for each location across the tibial plateaus and femoral condyles [28]. Prior to measuring cartilage thickness, outlier pixels disconnected from the main cartilage regions were manually removed and bone models were thoroughly reviewed. Cartilage thickness measurement reliability was evaluated by calculating a two-way, mixed effects, multiple-measurement, absolute agreement intraclass correlation coefficient for each cartilage surface. Additionally, we determined the coefficient of variation and standard deviation between the thickness measurements taken on different days.

### 2.3. Running-induced cartilage strain & recovery

Following testing of the bone and cartilage segmentation models, the trained models were used to perform the primary objective of this work, which was to measure tibial and femoral cartilage strain and recovery following a 3-mile run. In this analysis, all scans were manually cropped around the regions of interest, predicted by the trained segmentation models, reviewed, and compiled to generate 3D models of the bone and cartilage from which we could measure cartilage thickness. Obtaining thickness measurements at all three time points allowed for the comparison of cartilage thickness between the pre-exercise and recovery scans, as well as the calculation of cartilage strain post-exercise (Equation (1)).

$$Strain\ (\varepsilon) = 100 * \frac{Thickness_{post} - Thickness_{pre}}{Thickness_{pre}} \quad \text{Eq 1}$$

Two repeated measures ANOVAs were performed to determine the influence of time point (pre-exercise vs. post-exercise vs. recovery) on tibial and femoral cartilage thicknesses. Subsequent Fisher's Least Significant Difference post-hoc tests were performed to detect differences between means as appropriate. Differences were considered statistically significant with $p < 0.05$.

## 3. Results

### 3.1. Model development & testing

In terms of testing set DSCs, the 2D-UNet models outperformed the 3D-UNet models across all four tissues (Table 2). While the difference between the 2D- and 3D-UNet models was relatively small (Δ DSC <0.01), the difference was most pronounced for the cartilage models (Δ DSC for tibial cartilage = 0.006 and Δ DSC for femoral cartilage = 0.008). After optimal models were determined, they were subsequently applied to the testing set for each tissue (Table 2). Both bone models achieved testing set DSCs greater than 0.980, and both cartilage models achieved testing set DSCs greater than 0.900.

Evaluation of scan-to-scan repeatability demonstrated that bone surface distance group means were less than 0.03 mm for both the tibia and femur across all three directions (Table 3). Day-to-day measurements of cartilage thickness similarly revealed excellent reliability and precision, with ICCs of 0.984 and 0.987, and CVs of 1 % (standard deviation 0.03 mm) and 1 % (standard deviation 0.02 mm) for the tibial

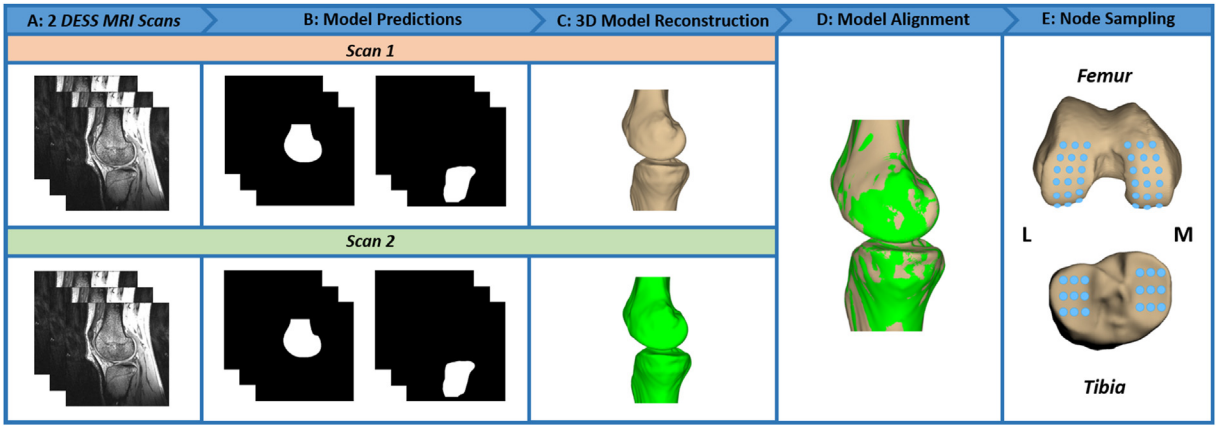


**Fig. 1. Bone Surface Distance Repeatability Analysis Pipeline. (A)** To assess bone surface repeatability, two DESS MRI scans from a single participant are inputted into the trained tibia and femur segmentation models. **(B)** The trained models predict the tibia and femur masks. **(C)** The outer contours from the predicted masks are extracted and reconstructed into 3D models using Geomagic Studio 11 (3D Systems; Research Triangle Park, North Carolina). **(D)** The tibia and femur models from Scan 1 and Scan 2 are aligned using an iterative closest point technique [6]. **(E)** Bone surface distances in the x-, y-, and z-directions are calculated between the two scans across all regions of interest [28]. M = medial and L = lateral.

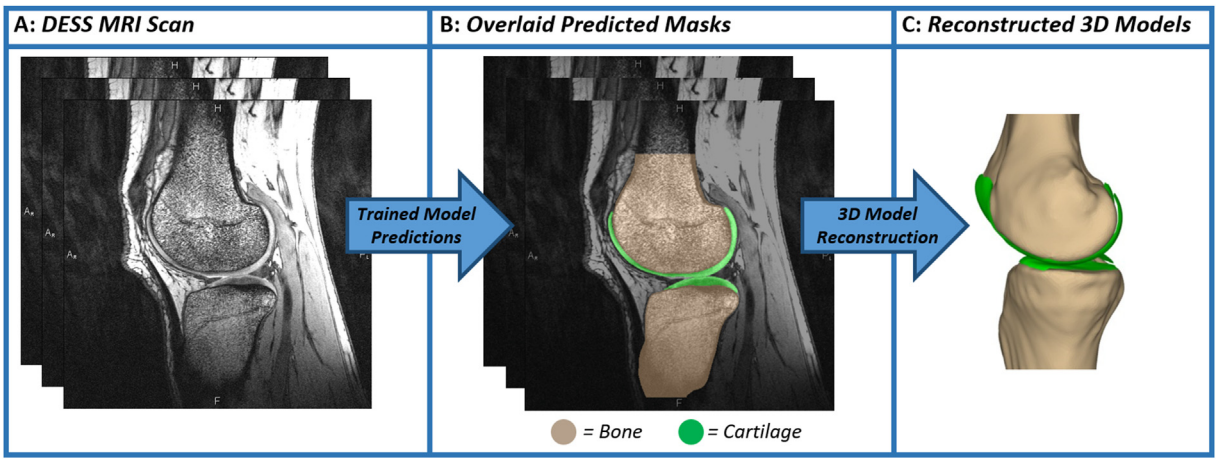


**Fig. 2. Bone and Cartilage 3D Model Reconstruction Pipeline. (A)** Each DESS MRI scan is inputted into the 4 trained segmentation models, and binary masks are outputted for each of the 4 tissues. **(B)** Visualization of the predicted bone and cartilage masks overlaid on the original DESS MRI scan. **(C)** Mask contours are extracted, converted into point clouds, and reconstructed into 3D surface mesh models using Geomagic Studio 11 (3D Systems; Research Triangle Park, North Carolina). Tan color indicates the tibia and femur, while green indicates the tibial and femoral cartilage.

**Table 2**
Dice Similarity Coefficient for bone and cartilage by dataset.

| Tissue | Validation Set | | Testing Set |
|---|---|---|---|
| | 2D-UNet | 3D-UNet | |
| Femoral bone | 0.989 | 0.987 | 0.990 |
| Tibial bone | 0.984 | 0.984 | 0.988 |
| Femoral cartilage | 0.921 | 0.913 | 0.913 |
| Tibial cartilage | 0.909 | 0.903 | 0.901 |

Testing set Dice Similarity Coefficients were calculated using the 2D-UNet.

**Table 3**
Mean (±95 % CI) scan-to-scan bone surface distances (mm).

| Direction | Tibia | Femur |
|---|---|---|
| X | 0.00 (0.04) | 0.01 (0.03) |
| Y | 0.02 (0.04) | 0.02 (0.03) |
| Z | 0.00 (0.03) | 0.01 (0.02) |

and femoral cartilage, respectively. Further, using our trained deep learning models, the total time required to segment the bone and cartilage in the tibiofemoral joint decreased from an order of hours or days when performed manually to an order of minutes when performed using deep learning.

### 3.2. Running-induced cartilage strain & recovery

In general, the 3-mile treadmill run induced compressive cartilage strains, and these strains were largely attenuated the morning following exercise (Fig. 3). An overall effect of time on cartilage thickness was observed for both the tibial ($p < 0.0001$) and femoral ($p < 0.01$) cartilage (Fig. 4). Tibial cartilage thickness changed by −0.15 (95 % CI = −0.20 to −0.11; $p < 0.0001$) mm pre-to post-exercise and recovered to 0.02 (95 % CI = −0.02 to 0.06; $p = 0.37$) mm the following day relative to baseline. Similarly, femoral cartilage thickness changed by −0.06 (95 % CI = −0.10 to −0.02; $p < 0.01$) mm pre-to post-exercise and recovered to 0.01 (95 % CI = −0.02 to 0.04; $p = 0.60$) mm the following day relative to baseline. Three miles of treadmill running resulted in compressive strains of 5.4 % (95 % CI = 4.1 to 6.7) and 2.3 % (95 % CI = 0.6 to 4.0) for the tibial and femoral cartilage, respectively (Fig. 4).

## 4. Discussion

This work sought to measure the deformation of tibiofemoral cartilage thickness immediately following a 3-mile treadmill run, as well as the recovery of cartilage thickness the following morning. In these analyses, both tibial and femoral cartilage thickness significantly decreased following the 3-mile run and recovered to within 1 % of pre-exercise thickness after one day of recovery. Additionally, deep learning models were trained to automate segmentation of the tibiofemoral bone and cartilage in order to enable the efficient measurement of cartilage thickness. All models were reliable and demonstrated precision similar to that of manual segmentation [28].

In terms of deep learning model development, we successfully trained models to segment the tibiofemoral bone and cartilage from DESS knee MRI scans. For each task, we trained both 2D- and 3D-UNet models to ensure that the optimal model architecture was chosen for each specific tissue. In terms of DSCs, the 2D models moderately outperformed the 3D models for each of the four tasks. While the primary benefit of the 3D-UNet architecture is the inclusion of inter-slice context, training of the 3D-UNet models was inherently constrained due to the computationally expensive nature of the 3D network and the relatively large image volumes. These factors led to a tradeoff between image volume size and maximum hyperparameter complexity due to memory constraints of the GPUs available for model training. Ultimately, the added inter-slice context of the 3D-UNet models may have not outweighed training restrictions, leading to the 2D models performing better. In terms of the optimal model testing set DSCs for each of the four tasks, our models compared favorably with the previously reported best scores [6,13,15,16,19,29–31]. It should be noted, however, that it is difficult to make direct comparisons between testing set DSCs across studies, as groups typically use different testing sets, which may impact the ultimate score. Despite this, the testing set DSCs from the present study reflect significant agreement between manually and automatically derived segmentations. One factor that likely contributed to our success in training models specific for our use-case was the consistency of data utilized for model development. All scans were collected on the same 3.0 T MRI scanner with identical imaging parameters [21–26], and all segmentations were reviewed by the same MSK radiologist with more than 35 years of experience. A limitation of our model development techniques was that training a specific model for each segmentation task now requires application of 4 separate models to predict the tissues of interest, whereas

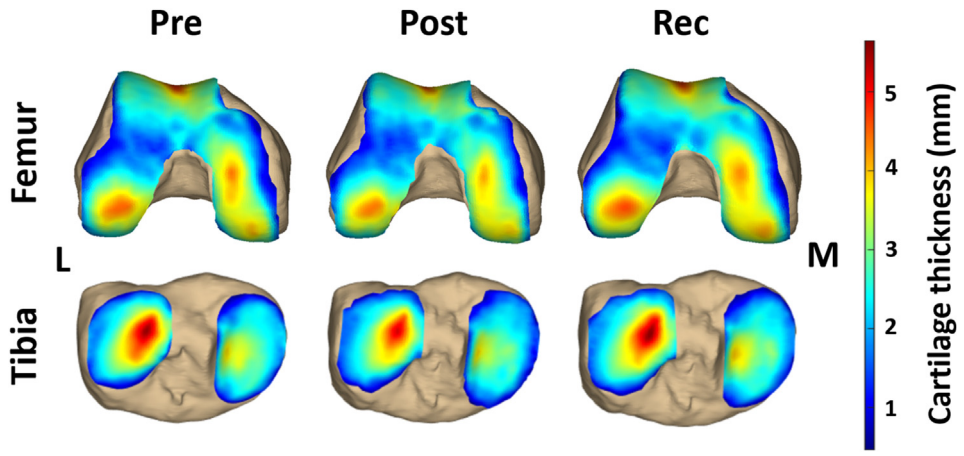


**Fig. 3. Cartilage thickness maps for one participant pre-exercise, post-exercise, and after 24-h of recovery.** Red represents thicker regions of cartilage, while blue represents thinner regions. Post-exercise cartilage thickness decreased compared to the pre-exercise and recovery time points, which were relatively similar. M = medial and L = lateral.

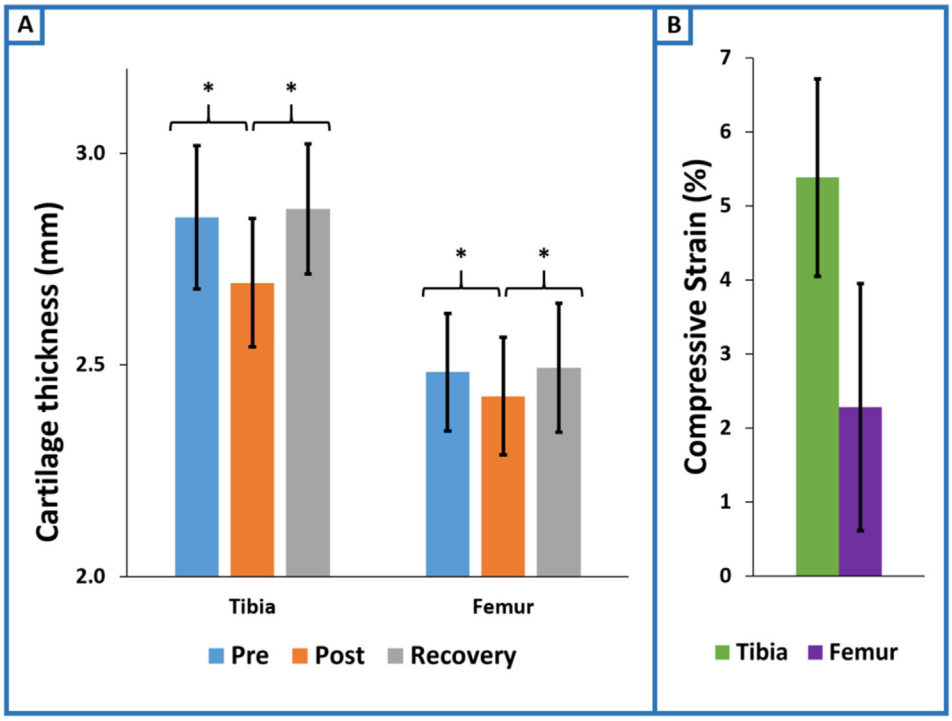


**Fig. 4. Running-induced changes to cartilage thickness and strain. (A)** Cartilage thickness decreased significantly pre-to post-exercise for both the tibial ($p < 0.0001$) and femoral ($p < 0.01$) cartilage. Subsequently, cartilage thickness increased back towards baseline thickness at the recovery time point for both the tibial ($p < 0.0001$) and femoral ($p < 0.01$) cartilage. Significant differences between baseline and recovery were not detected for either the tibial ($p = 0.37$) or femoral ($p = 0.60$) cartilage. **(B)** Running induced post-exercise compressive strains of 5.4 % (95 % CI = 4.1 to 6.7) and 2.3 % (95 % CI = 0.6 to 4.0) for the tibial and femoral cartilage, respectively.

a multiclass model could have predicted each tissue with a single model. However, training a separate model for each segmentation task enabled task-specific hyperparameter optimization. This also enabled cropping of the image volume to each tissue, which decreased computational cost and ultimately enabled more hyperparameter combinations to be tested during model training. While the objective of this work was to train models to segment the tibiofemoral bone and cartilage from DESS knee MRI scans with specific imaging parameters, a limitation is therefore that the application of these models is thus specific in nature. Specifically, it is not well known how they would perform with imaging data that differs from that which the models were trained on.

Following model development, we evaluated the repeatability of using the trained segmentation models to create 3D models of the tibia and femur, as well as measure cartilage thickness. Beginning with the tibia and femur, we utilized two scans per participant that were obtained approximately 30 min apart after the participant was repositioned and measured scan-to-scan group mean distances less than 0.03 mm for the x-, y-, and z-directions. While some distances represent a small positive bias, the 95 % confidence intervals of all distances overlap with zero. Further, while these distances should largely represent differences in scan-to-scan segmentation, they may also encompass some error introduced from 3D model alignment. However, most importantly, the bone surface distances measured here all fall within the previously reported manual measurement resolution for this technique [28], meaning that the repeatability of using the trained bone models is comparable to that of trained manual segmenters plus review from a fellowship-trained radiologist. The repeatability of creating bone 3D models is important, as it has been suggested that bone shape may change over time with injury, aging, and OA status [32–36]. Next, we evaluated the repeatability of using the bone and cartilage segmentation models to measure cartilage thickness. A limitation of this work is that we did not perform manual segmentations to compare manual and automated strain

measurements. However, in order to assess measurement reliability, we measured cartilage thickness in the same cohort of participants across two different days. Assessing measurement reliability across different days in this way has an advantage over comparison to manual segmentations as it removes intra- and inter-segmenter variability that may bias the perceived validity of the model. The ICCs for this analysis were greater than 0.980 for both the tibial and femoral cartilage, indicating excellent measurement reliability [37]. Further, the day-to-day measurement of cartilage thickness resulted in CVs of 1 % for both the tibial and femoral cartilage. The present results indicate that our resolution for measuring cartilage thickness is similar to the previously established resolution of this technique using manual segmentation (~1 %) [28].

The primary objective of this work was to use our trained models to measure tibial and femoral cartilage strain and recovery in response to a 3-mile treadmill run. Prior work has investigated running-induced changes to cartilage thickness and volume with varying results [7,38–42]. While several studies have demonstrated changes in cartilage thickness and volume in some compartments after running [38,40–42], others have failed to detect significant changes in tibiofemoral cartilage thickness [39]. In the present study, using site-specific measurements of cartilage thickness, we detected a significant decrease in both the tibial (5.4 %) and femoral (2.3 %) cartilage (Fig. 4). The tibial and femoral cartilage strain magnitudes in this study align with prior work that reported higher strains in the tibial cartilage compared to the femoral cartilage in response to walking and hopping [21,22,43]. Potential explanations for why femoral cartilage experiences less compressive strain include differences in the mechanical properties and loading between the femur and tibia [22,44]. Specifically, prior work has reported that femoral cartilage is stiffer and less permeable compared to tibial cartilage [44,45]. With regard to loading differences, as the knee flexes during dynamic activities, the regions of the femoral cartilage in contact change more than those of the tibial cartilage [22,43]. These differences may cause the loading to be more dispersed, reducing strain on the femoral cartilage [22,43].

Prior work has also utilized quantitative MRI techniques, such as T1rho and T2 relaxation mapping to evaluate changes in cartilage composition following loading [20,38–40,46]. For example, a recent study from our lab found that running results in decreased tibial, femoral, and patellar cartilage T1rho relaxation times, indicative of an increase in the concentration of proteoglycans due to expulsion of water [20]. Similarly, decreased T2 relaxation times have been measured in the superficial regions of tibiofemoral cartilage immediately following running, likely resulting from a reduction in water content [46]. Thus, our present findings support the hypothesis that running results in immediate post-exercise changes in both tibial and femoral cartilage thickness due to water loss from the cartilage.

In addition to probing the immediate influence of a 3-mile run, tibiofemoral cartilage thickness was also measured the following morning to elucidate the ability of cartilage to recover at a time scale relevant to habitual exercise. At the recovery time point, both the tibial and femoral cartilage thickness had significantly increased from the post-exercise thickness, and recovered to within 1 % of baseline. With regard to prolonged effects after exercise, several prior studies have probed cartilage thickness and composition between 2 and 12 h following a marathon run [47–49]. However, in these studies, cartilage thickness and composition were not measured immediately following exercise, making it difficult to draw direct comparisons to the present study. Prior work from our lab evaluated tibial, femoral, and patellar cartilage T1rho relaxation times before, immediately following, and approximately 24 h following a 3 or 10 mile run [20]. T1rho relaxation times were significantly decreased immediately post-exercise, and returned to pre-exercise levels after approximately 24 h of recovery [20]. Our present findings, in conjunction with these prior results, suggest that tibiofemoral cartilage thickness and composition have the ability to recover from a 3-mile run within 24 h. However, a limitation of this work is that the participants were relatively young and healthy. Future work is needed to investigate how repetitive bouts of running may alter cartilage strain and recovery as it relates to long-term joint health, in addition to evaluating the influence of age, sex, BMI, and knee health status.

In conclusion, we measured tibiofemoral cartilage strain and recovery following a 3-mile treadmill run, enabled via the development of deep learning models to automate segmentation of the tibiofemoral cartilage and bone. In this analysis, we demonstrated that a 3-mile run induced both tibial and femoral cartilage compressive strains, and after approximately one day, the cartilage had largely returned to baseline thickness. This suggests that a single 3-mile run may be a safe magnitude of loading for tibial and femoral cartilage in a relatively young and healthy cohort of participants. Future work will seek to expand upon these findings in a more diverse participant population and under varying run distances and intensities.

## Author contributions

**PXB:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original draft preparation, Writing – Reviewing and editing; **SYK:** Methodology, Software, Data curation, Writing – Reviewing and editing; **BSB:** Validation, Formal analysis, Investigation, Writing – Reviewing and editing; **ADR:** Conceptualization, Methodology, Investigation, Data curation, Writing – Reviewing and editing; **ATC:** Investigation, Data curation, Writing – Reviewing and editing; **LNH:** Conceptualization, Methodology, Investigation, Data curation, Writing – Reviewing and editing; **ECO:** Investigation, Data curation, Writing – Reviewing and editing; **MRW:** Investigation, Data curation, Writing – Reviewing and editing; **CSP:** Investigation, Data curation, Writing – Reviewing and editing; **BSC:** Investigation, Data curation, Writing – Reviewing and editing; **NKL:** Investigation, Data curation, Writing – Reviewing and editing; **EGS:** Investigation, Data curation, Writing – Reviewing and editing; **BPM:** Conceptualization, Methodology, Writing – Reviewing and editing; **CES:** Conceptualization, Methodology, Writing – Reviewing and editing; **LED:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Investigation, Writing – Reviewing and editing.

## Conflict of interest

This work was prepared while Dr. Amber Collins was employed at Duke University School of Medicine. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government. The remaining authors have no conflicts of interest to disclose.

## Role of the funding source

This work was supported by NIH grants R01AR074800, R01AR065527, R01AR082665, and R01AR079184.

## Acknowledgements

Writing support was provided by Donald T. Kirkendall.

## References

[1] D.D. Chan, L. Cai, K.D. Butz, S.B. Trippel, E.A. Nauman, C.P. Neu, *In vivo* articular cartilage deformation: noninvasive quantification of intratissue strain during joint contact in the human knee, Scientific Reports 6 (2016) 19220.
[2] X. Li, C. Benjamin Ma, T.M. Link, D.D. Castillo, G. Blumenkrantz, J. Lozano, et al., *In vivo* T1ρ and T2 mapping of articular cartilage in osteoarthritis of the knee using 3T MRI, Osteoarthritis Cartilage 15 (2007) 789–797.
[3] F. Eckstein, A.E. Wluka, W. Wirth, F. Cicuttini, 30 Years of MRI-based cartilage & bone morphometry in knee osteoarthritis: from correlation to clinical trials, Osteoarthritis Cartilage 32 (2024) 439–451.
[4] F. Guilak, Biomechanical factors in osteoarthritis, Best Practice & Research Clinical Rheumatology 25 (2011) 815–823.

[5] S.L. Coburn, K.M. Crossley, J.L. Kemp, S.J. Warden, T.J. West, A.M. Bruder, et al., Is running good or bad for your knees? A systematic review and meta-analysis of cartilage morphology and composition changes in the tibiofemoral and patellofemoral joints, Osteoarthritis Cartilage 31 (2023) 144–157.
[6] S. Khan, B. Azam, Y. Yao, W. Chen, Deep collaborative network with alpha matte for precise knee tissue segmentation from MRI, Comput Methods Programs Biomed 222 (2022) 106963.
[7] L.N. Heckelman, A.D. Riofrio, E.N. Vinson, A.T. Collins, O.R. Gwynn, G.M. Utturkar, et al., Dose and recovery response of patellofemoral cartilage deformations to running, Orthopaedic Journal of Sports Medicine 8 (2020) 2325967120967512.
[8] E.C. Brenneman Wilson, A.A. Gatti, M.R. Maly, A new technique to evaluate the impact of running on knee cartilage deformation by region, Magnetic Resonance Materials in Physics, Biology and Medicine 34 (2021) 593–603.
[9] F. Eckstein, B. Lemberger, C. Gratzke, M. Hudelmaier, C. Glaser, K.H. Englmeier, et al., *In vivo* cartilage deformation after different types of activity and its dependence on physical training status, Annals of the Rheumatic Diseases 64 (2005) 291–295.
[10] J.A. Coppock, N.E. Zimmer, C.E. Spritzer, A.P. Goode, L.E. DeFrate, Automated segmentation and prediction of intervertebral disc morphology and uniaxial deformations from MRI, Osteoarthritis and Cartilage Open 5 (2023) 100378.
[11] L.N. Heckelman, B.J. Soher, C.E. Spritzer, B.D. Lewis, L.E. DeFrate, Design and validation of a semi-automatic bone segmentation algorithm from MRI to improve research efficiency, Scientific Reports 12 (2022) 7825.
[12] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.
[13] S.Y. Kim-Wang, P.X. Bradley, H.C. Cutcliffe, A.T. Collins, B.S. Crook, C.S. Paranjape, et al., Auto-segmentation of the tibia and femur from knee MR images via deep learning and its application to cartilage strain and recovery, Journal of Biomechanics 149 (2023) 111473.
[14] S.W. Flannery, A.M. Kiapour, D.J. Edgar, M.M. Murray, B.C. Fleming, Automated magnetic resonance image segmentation of the anterior cruciate ligament, Journal of Orthopaedic Research 39 (2021) 831–840.
[15] S. Gaj, M. Yang, K. Nakamura, X. Li, Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks, Magnetic Resonance in Medicine 84 (2019) 437–449.
[16] B. Norman, V. Pedoia, S. Majumdar, Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry, Radiology 288 (2018) 177–185.
[17] C.P.S. Kulseng, V. Nainamalai, E. Grøvik, J.T. Geitung, A. Årøen, K.I. Gjesdal, Automatic segmentation of human knee anatomy by a convolutional neural network applying a 3D MRI protocol, BMC Musculoskelet Disorder 24 (2023) 41.
[18] A.D. Desai, F. Caliva, C. Iriondo, A. Mortazi, S. Jambawalikar, U. Bagci, et al., The international workshop on osteoarthritis imaging knee MRI segmentation challenge: a multi-institute evaluation and analysis framework on a standardized dataset, Radiology: Artificial Intelligence 3 (2021) e200078.
[19] A.A. Gatti, M.R. Maly, Automatic knee cartilage and bone segmentation using multi-stage convolutional neural networks: data from the osteoarthritis initiative, Magma 34 (2021) 859–875.
[20] L.N. Heckelman, W.A.R. Smith, A.D. Riofrio, E.N. Vinson, A.T. Collins, O.R. Gwynn, et al., Quantifying the biochemical state of knee cartilage in response to running using T1rho magnetic resonance imaging, Scientific Reports 10 (2020) 1870.
[21] B.S. Crook, A.T. Collins, N.K. Lad, C.E. Spritzer, J.R. Wittstein, L.E. DeFrate, Effect of walking on *in vivo* tibiofemoral cartilage strain in ACL-deficient versus intact knees, Journal of Biomechanics 116 (2021) 110210.
[22] N.K. Lad, B. Liu, P.K. Ganapathy, G.M. Utturkar, E.G. Sutter, C.T. Moorman, et al., Effect of normal gait on *in vivo* tibiofemoral cartilage strains, Journal of Biomechanics 49 (2016) 2870–2876.
[23] E.C. Okafor, G.M. Utturkar, M.R. Widmyer, E.S. Abebe, A.T. Collins, D.C. Taylor, et al., The effects of femoral graft placement on cartilage thickness after anterior cruciate ligament reconstruction, Journal of Biomechanics 47 (2014) 96–101.
[24] C.S. Paranjape, H.C. Cutcliffe, S.C. Grambow, G.M. Utturkar, A.T. Collins, W.E. Garrett, et al., A new stress test for knee joint cartilage, Scientific Reports 9 (2019) 2283.
[25] E.G. Sutter, B. Liu, G.M. Utturkar, M.R. Widmyer, C.E. Spritzer, H.C. Cutcliffe, et al., Effects of anterior cruciate ligament deficiency on tibiofemoral cartilage thickness and strains in response to hopping, The American Journal of Sports Medicine 47 (2019) 96–103.
[26] M.R. Widmyer, G.M. Utturkar, H.A. Leddy, J.L. Coleman, C.E. Spritzer, I.I.I.C.T. Moorman, et al., High body mass index is associated with increased diurnal strains in the articular cartilage of the knee, Arthritis & Rheumatism 65 (2013) 2615–2622.
[27] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, et al., Statistical validation of image segmentation quality based on a spatial overlap index, Academic Radiology 11 (2004) 178–189.
[28] J.L. Coleman, M.R. Widmyer, H.A. Leddy, G.M. Utturkar, C.E. Spritzer, C.T. Moorman 3rd, et al., Diurnal variations in articular cartilage thickness and strain in the human knee, Journal of Biomechanics 46 (2013) 541–547.
[29] F. Ambellan, A. Tack, M. Ehlke, S. Zachow, Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the Osteoarthritis Initiative, Medical Image Analysis 52 (2019) 109–118.
[30] M.H.A. Latif, I. Faye, Automated tibiofemoral joint segmentation based on deeply supervised 2D-3D ensemble U-Net: data from the Osteoarthritis Initiative, Artificial Intelligence in Medicine 122 (2021) 102213.
[31] E. Panfilov, A. Tiulpin, S. Klein, M.T. Nieminen, S.S. Saarakkala, Improving robustness of deep learning based knee MRI segmentation: mixup and adversarial domain adaptation, in: IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 450–459, 2019.
[32] V. Pedoia, D.A. Lansdown, M. Zaid, C.E. McCulloch, R. Souza, C.B. Ma, et al., Three-dimensional MRI-based statistical shape model and application to a cohort of knees with acute ACL injury, Osteoarthritis Cartilage 23 (2015) 1695–1703.
[33] D.J. Haverkamp, D. Schiphof, S.M. Bierma-Zeinstra, H. Weinans, J.H. Waarsing, Variation in joint shape of osteoarthritic knees, Arthritis & Rheumatism 63 (2011) 3401–3407.
[34] C. Ding, F. Cicuttini, F. Scott, H. Cooley, G. Jones, Association between age and knee structural change: a cross sectional MRI based study, Annals of the Rheumatic Diseases 64 (2005) 549–555.
[35] M.A. Bowes, R.A. Maciewicz, J.C. Waterton, D.J. Hunter, P.G. Conaghan, Bone area provides a responsive outcome measure for bone changes in short-term knee osteoarthritis studies, Journal of Rheumatology 43 (2016) 2179–2182.
[36] T.L. Bredbenner, T.D. Eliason, R.S. Potter, R.L. Mason, L.M. Havill, D.P. Nicolella, Statistical shape modeling describes variation in tibia and femur surface geometry between Control and Incidence groups from the osteoarthritis initiative database, Journal of Biomechanics 43 (2010) 1780–1786.
[37] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, Chiropractic Journal of Medicine 15 (2016) 155–163.
[38] A.A. Gatti, M.D. Noseworthy, P.W. Stratford, E.C. Brenneman, S. Totterman, J. Tamez-Peña, et al., Acute changes in knee cartilage transverse relaxation time after running and bicycling, Journal of Biomechanics 53 (2017) 171–177.
[39] K. Subburaj, D. Kumar, R.B. Souza, H. Alizai, X. Li, T.M. Link, et al., The acute effect of running on knee articular cartilage and meniscus magnetic resonance relaxation times in young healthy adults, The American Journal of Sports Medicine 40 (2012) 2134–2141.
[40] E.C. Brenneman Wilson, A.A. Gatti, P.J. Keir, M.R. Maly, Daily cumulative load and body mass index alter knee cartilage response to running in women, Gait Posture 88 (2021) 192–197.
[41] A. Van Ginckel, P. Verdonk, J. Victor, E. Witvrouw, Cartilage status in relation to return to sports after anterior cruciate ligament reconstruction, The American Journal of Sports Medicine 41 (2013) 550–559.
[42] M. Boocock, P. McNair, F. Cicuttini, A. Stuart, T. Sinclair, The short-term effects of running on the deformation of knee articular cartilage and its relationship to biomechanical loads at the knee, Osteoarthritis and Cartilage 17 (2009) 883–890.
[43] E.G. Sutter, M.R. Widmyer, G.M. Utturkar, C.E. Spritzer, W.E. Garrett Jr., L.E. DeFrate, *In vivo* measurement of localized tibiofemoral cartilage strains in response to dynamic activity, The American Journal of Sports Medicine 43 (2015) 370–376.
[44] S. Treppo, H. Koepp, E.C. Quan, A.A. Cole, K.E. Kuettner, A.J. Grodzinsky, Comparison of biomechanical and biochemical properties of cartilage from human knee and ankle pairs, Journal of Orthopaedic Research 18 (2000) 739–748.
[45] B.L. Wong, R.L. Sah, Mechanical asymmetry during articulation of tibial and femoral cartilages: local and overall compressive and shear deformation and properties, Journal of Biomechanics 43 (2010) 1689–1695.
[46] H.A. Crowder, V. Mazzoli, M.S. Black, L.E. Watkins, F. Kogan, B.A. Hargreaves, et al., Characterizing the transient response of knee cartilage to running: decreases in cartilage T(2) of female recreational runners, Journal of Orthopaedic Research 39 (2021) 2340–2352.
[47] M. Zhang, Y. Li, R. Feng, Z. Wang, W. Wang, N. Zheng, et al., Change in susceptibility values in knee cartilage after marathon running measured using quantitative susceptibility mapping, Journal of Magnetic Resonance Imaging 54 (2021) 1585–1593.
[48] P. Zhang, B. Yu, R. Zhang, X. Chen, S. Shao, Y. Zeng, et al., Longitudinal study of the morphological and T2* changes of knee cartilages of marathon runners using prototype software for automatic cartilage segmentation, British Journal of Radiology 94 (2021) 20200833.
[49] Z. Wang, S. Ai, F. Tian, M.H.L. Liow, S. Wang, J. Zhao, et al., Higher body mass index is associated with biochemical changes in knee articular cartilage after marathon running: a quantitative T2-relaxation MRI study, Orthopaedics Journal of Sports Medicine 8 (2020) 2325967120943874.